



# European Data Initiatives

October 2, 2012

Leif Laaksonen, Peter Wittenburg  
iCordi

# Agenda for today

- ✧ A process picture of the European data initiatives
- ✧ The European angle into Research Data Alliance through the iCORDI project

# Data initiatives in Europe

- ✧ There is already a long history of data management projects and initiatives in Europe, with several existing data infrastructures dealing with established and growing user communities:
  - ✧ European Center for Nuclear Research (CERN),
  - ✧ European Space Agency (ESA),
  - ✧ European Organisation for Astronomical Research (ESO),
  - ✧ European Bioinformatics Institute (EBI)
  - ✧ EIROforum is a partnership between eight of Europe's largest inter-governmental scientific research organisations that are responsible for infrastructures and laboratories: CERN, EFDA-JET, EMBL, ESA, ESO, ESRF, European XFEL and ILL.
  - ✧ The numerous Research Infrastructures being created as part of the **European Strategy Forum on Research Infrastructures (ESFRI)** roadmap can also be seen as data management initiatives, connecting repositories, aggregating and sharing data across national borders, and developing tools to make these data widely available to their communities and beyond. Latest update 2010.

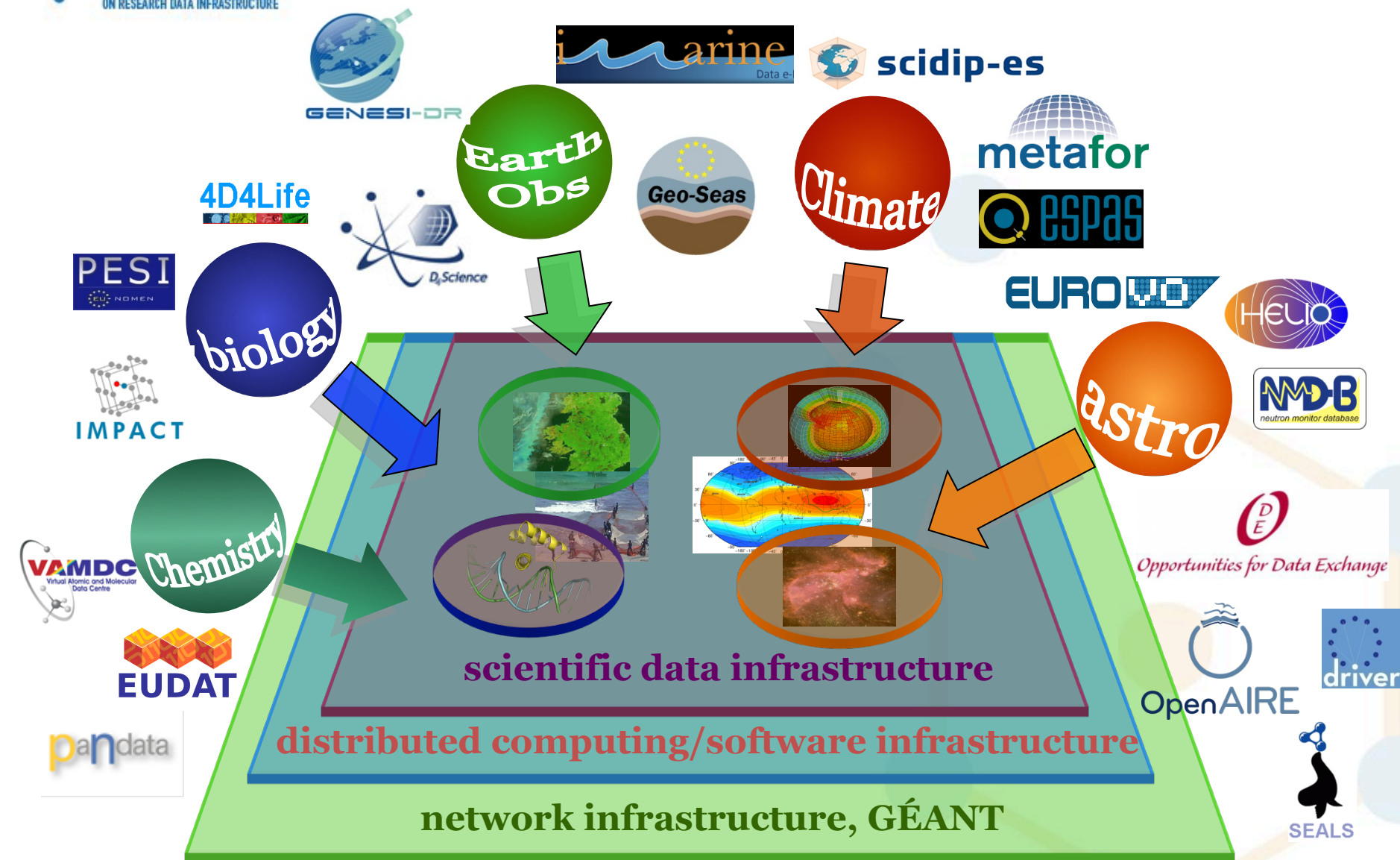
# Areas of ESFRI Roadmap 2010

- ✧ Social Sciences and Humanities
- ✧ Environmental Sciences
- ✧ Energy
- ✧ Biological and Medical Sciences now called Health and Food Domain
- ✧ Materials and Analytical Facilities
- ✧ Physical Sciences and Engineering
- ✧ e-Infrastructures (PRACE)

# Clustering of ESFRI RI

- ✧ [DASISH](#) provides solutions to a number of common issues for the five projects in social sciences and humanities: CESSDA, CLARIN, DARIAH, ESS and SHARE;
- ✧ The [BioMedBridges](#) consortium brings together the six under implementation ESFRI infrastructures (ELIXIR, BBMRI, EATRIS, ECRIN, InfraFrontier and INSTRUCT);
- ✧ The [ENVRI project](#) gathers seven ESFRI Environment projects (the others are represented in the advisory board) to develop, with the support from ICT experts, common data and software components and services for their facilities;
- ✧ The CRISP proposal focuses on four R&D tasks that are of utmost importance for these RIs: (i) Accelerators, (ii) Instruments & Experiments, (iii) Detectors & Data Acquisition, and (iv) Information Technology (IT) & Data Management;

# 100 million euro EC investment



# EC data FP7 initiatives (1/3)

- 🏠 OpenAIRE and OpenAIREplus provide an electronic infrastructure supporting mechanisms for the identification, deposit, access, and monitoring of the European programmes;
- 🏠 EUDAT framework for multidisciplinary data e-infrastructure filling the gaps between computing (including HPC) and data driven applications;
- 🏠 EuroVO-AIDA: developing the EURO Virtual Observatory (VO) deploying in Europe an operational technical infrastructure to support shared access to astronomic data ;
- 🏠 GEOSEAS: interoperability of datasets and improvement on access to marine geological and geophysical data and data-products from national geological surveys and research institutes in Europe;
- 🏠 iMARINE open and collaborative initiative of a data infrastructure to support fisheries management and conservation of marine living resources;
- 🏠 4D4Life: on biodiversity to strengthen the development of GlobalSpecies Databases extending the geographical reach by realizing a Multi-Hub Network integrating data from China, New Zealand, Australia, N. America and Brazil;



## EC data FP7 initiatives (2/3)

- ✧ ESPAS e-Infrastructure to support the access to observations, the modelling and prediction of the near-Earth space environment;
- ✧ METAFOR: on databases used in Climate Change;
- ✧ VAMDC building a secure, flexible and interoperable e-science environment interface to the existing Atomic and Molecular data;
- ✧ ODE (Opportunities for Data Exchange, led by CERN; ESA and the Association for Permanent Access);
- ✧ SCIDIP-ES (European Space Agency led Long Term Data Preservation programme;
- ✧ IMPACT led by the European Molecular and Biological Lab on global interoperability and curation of protein databases etc;



# EC data FP7 initiatives (3/3)

- ✧ COOPEUS: Building a Framework for a Sustainable, Transatlantic Cooperation in the Field of Environmental Research Infrastructures
- ✧ ODIP: Establishing and operating an Ocean Data Interoperability Platform
- ✧ DRIHM2US: Distributed Research Infrastructure for Hydro-Meteorology to United States of America
- ✧ **Infrastructures that produce significant data amounts:**
  - ✧ Partnership for Advanced Computing in Europe (PRACE), Brussels
  - ✧ European Grid Initiative (EGI), Amsterdam



# The iCordi project

# iCORDI key facts

Project Name	iCORDI – Global Data Interoperability
Start date	1st September 2012
Duration	24 months
Budget	3,3 M€ (including 2,3 M€ from the EC) and 240 PMs
EC call	Call 10 (INFRA-2012-3.2): International cooperation with the USA on common e-Infrastructure for scientific data (11.2011)
Participants	14 partners from 8 countries (national data centers, technology providers and research communities)
Objectives	“iCORDI’s prime objective is to establish a <b>coordination platform</b> between Europe and the USA to discuss and improve the interoperability of today’s and tomorrow’s scientific data infrastructures of both continents and to extend this to the global levels.”

# iCORDI goals

- ✦ Facilitate and support the creation of an IETF (Internet Engineering Task Force) like organisation for the global data interoperability and data exchange – **Research Data Alliance (RDA)**
- ✦ Support the work of European research communities in their effort to collaborate with their US counter parts through a **coordinated platform** and to extend this to the global levels

# Total iCORDI effort 240PM

- ✧ WP1 (Project management) 14PM
  - ✧ T1.1 Management (14PM)
- ✧ WP2 (Access and interoperability platform) 98PM
  - ✧ T2.1 Supporting the Research Data Alliance (RDA) process (20PM)
  - ✧ T2.2 Setting up and supporting the High Level Scientific/Strategic Forum (HLSF) as stimulation activity (29PM)
  - ✧ T2.3 Data organization analysis and support (49PM)
- ✧ WP3 (Community-led collaboration) 90PM
  - ✧ T3.1 Community Workshop Programme (40PM)
  - ✧ T3.2 Community Prototype Programme (50PM)
- ✧ WP4 (Dissemination and event support) 38PM
  - ✧ T4.1 Dissemination activities (18PM)
  - ✧ T4.2 Events & workshops (20PM)

# A coordinated platform (1/2)

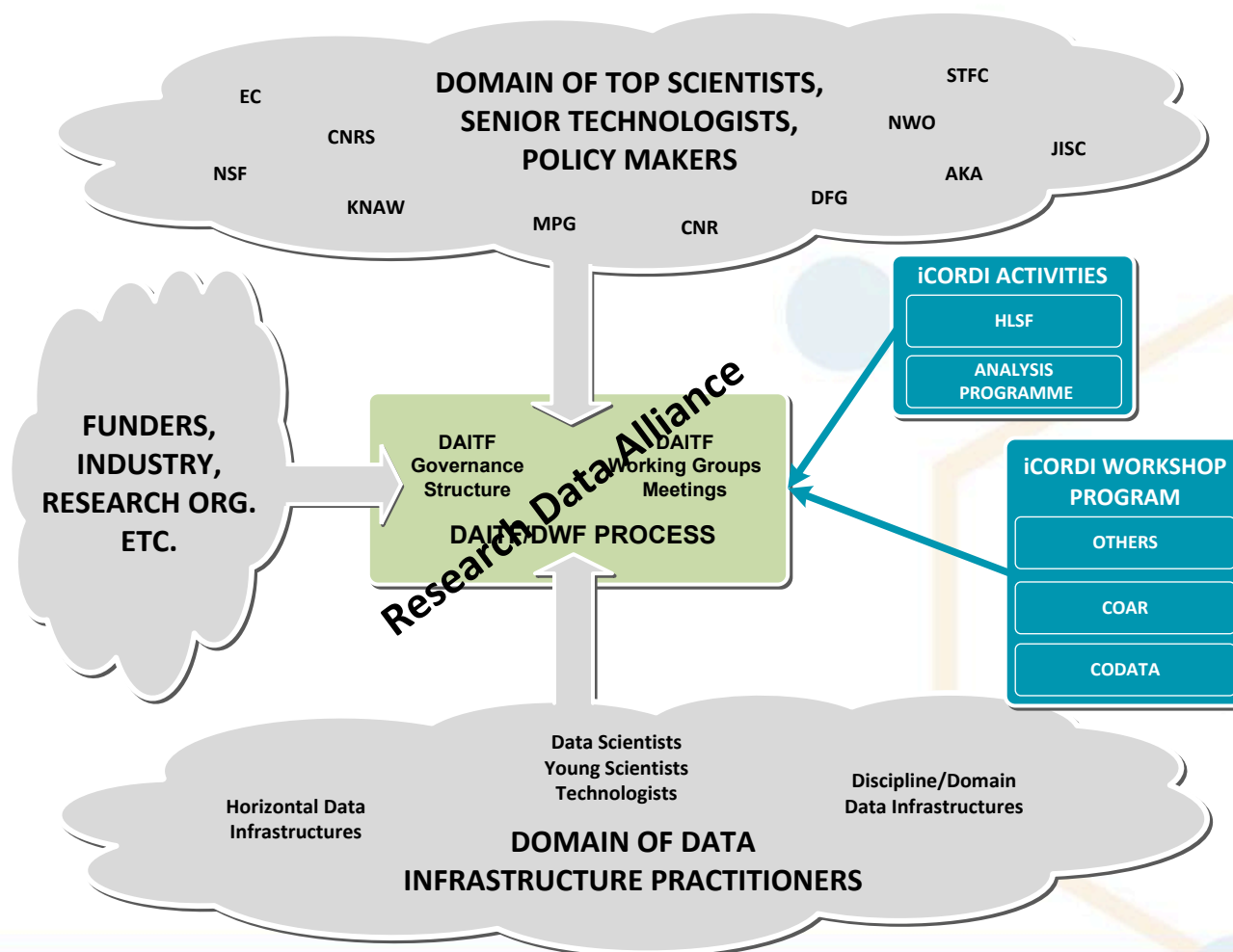
- ✧ By a coordination platform, we mean a vehicle made of several complementary components and capable of:
  - ✧ fostering discussion between relevant stakeholders in particular in the EU and US/global over concrete topics related to the interoperability of the data architectures and solutions based on a top-down approach;
  - ✧ overcoming the identified challenges and turning the areas of convergence into concrete specifications that can be immediately implemented on both continents by bringing data practitioners together in a bottom-up process;
  - ✧ demonstrating through concrete examples of collaboration what works and what are the remaining barriers and challenges to be tackled to achieve full interoperability.

# A coordinated platform (2/2)

- ✧ The platform that iCORDI will set up will be supported by three key programmes:
  - ✧ An *analysis programme* devoted to analyzing data organizations and solutions as they emerge from the various scientific communities.
  - ✧ A *prototype programme* to coordinate a series of cross-infrastructure experiments on EU-US interoperability with a selected group of important projects and communities.
  - ✧ A *workshop programme* covering the investigation data infrastructure convergence and reaching out to new research communities.



# iCORDI ecosystem



# iCORDI activities (1/2)

- Supporting the emerging Research Data Alliance (RDA) grass-root process as part of the international collaboration platform by helping to prepare the format and to define the contents of the RDA meetings and by helping to organise the working groups.
- Supporting RDA and the High Level Strategy Forum (HLSF), which is now European, with an analysis programme that focuses on cross-disciplinary and cross-national comparisons of solutions and formulates possible abstractions in order to establish best practices.

## iCORDI activities (2/2)

- ✧ To complement and support the work of the High Level Strategy Forum (HSLF) with community-led programmes of knowledge exchange and prototyping.
- ✧ To capture and collate results and feed genuine experimental outcomes, requirements and issues into the RDA and HSLF discussion and analysis processes.

# Contact contact details

- ✧ Project Coordinator: Leif Laaksonen  
([leif.laaksonen@csc.fi](mailto:leif.laaksonen@csc.fi))
- ✧ Scientific Coordinator: Peter Wittenburg  
([peter.wittenburg@mpi.nl](mailto:peter.wittenburg@mpi.nl))
- ✧ Contact: [icordi-info@postit.csc.fi](mailto:icordi-info@postit.csc.fi)
- ✧ Web site: [www.icordi.eu](http://www.icordi.eu)

# Who's Involved in iCORDI



[icordi-info@postit.csc.fi](mailto:icordi-info@postit.csc.fi) | [www.icordi.eu](http://www.icordi.eu)